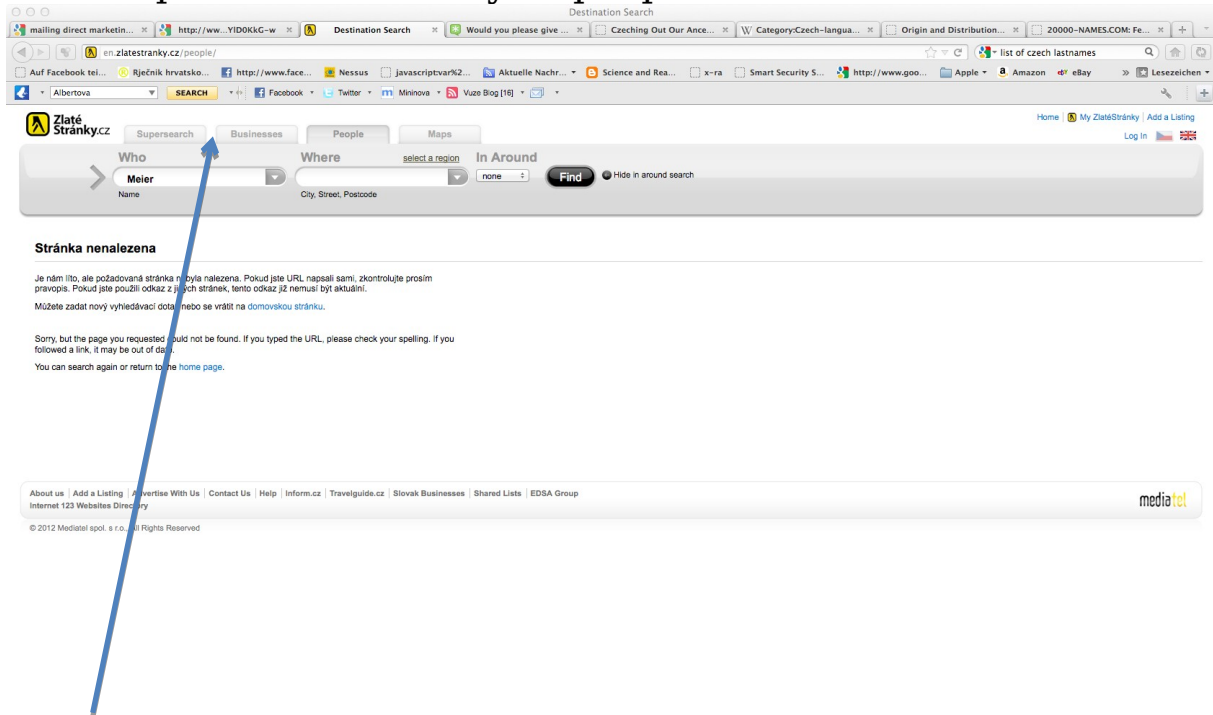


Addresscollector from Zlate Strank in Czech Republic

The script should crawl all possible combinations of names on the webpage of Zlatestranky:

<http://en.zlatestranky.cz/people/>



Here is the lastname and search parameter entered as well as the city...

You can find a list of all Czech Surnames here:

http://www.iarerelative.com/czech/a_b.htm

<http://www.iarerelative.com/czech/alpha.htm>

http://en.wikipedia.org/wiki/Category:Czech-language_surnames

You should write the script like that, that you check all lastnames in every ZIP code, where you check in 10000, 11000, 12000, 13000, ...till , 99000 where you add the extension that you search in an area of 75 kilometer:

<http://en.zlatestranky.cz/people/12000/Meier/1/?distance=75>

The search should be made like this:

The screenshot shows the search interface of the website en.zlatestranky.cz/people/12000/Meier/1/?distance=75. The interface includes tabs for 'Supersearch', 'Businesses', 'People', and 'Maps'. The 'Who' section has a search bar with 'Meier' entered, labeled 'Name' below it. The 'Where' section has a search bar with '120 00' entered, labeled 'City, Street, Postcode' below it. The 'In Around' section has a dropdown menu set to '<75 km'. A 'Find' button is located to the right of the 'In Around' section. Three blue arrows point from text labels to the search fields: 'Lastname' points to the 'Who' field, 'Zip Code' points to the 'Where' field, and 'Add 75' points to the 'In Around' dropdown.

Who: Meier (Name)

Where: 120 00 (City, Street, Postcode)

In Around: <75 km

Find

Lastname

Zip Code

Add 75

The results will look like that:



When you click on an entry you see it like that
and you get an option to download the record
as VCF file (Outlook Address File)

More filter options: [All](#) [Places](#) [Name](#) Sorted by: [Relevance](#) [Distance](#)

A Meier Daniel
Kamenická 652/48, 17000 Praha 7 - Holešovice
233 373 975

B Meier Erich
Kamenická 652/48, 17000 Praha - Holešovice
222 954 878

C Meier Ervin
Družstevní 24, 27351 Braškov
312 699 313

D Meier Manfred
K Lesu 1350, 25301 Hostivice
220 980 982

< Back Map Get Directions Save to: [My Places](#) Send to: [Outlook](#) [Email](#)

Meier Daniel
Kamenická 652/48
17000 Praha 7 - Holešovice
Tel.:233 373 975

[See large map](#)

When you download the adress file it will be a text file looking like that:

```
BEGIN:VCARD
VERSION:2.1

N;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:Meier Daniel
FN;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:Meier Daniel
TEL;TYPE=WORK;TYPE=VOICE:233 373 975
ADR;WORK;ENCODING=QUOTED-PRINTABLE;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-
CZ;CHARSET=ISO-8859-2;;Kamenick=E1 652/48;Praha;Hlavn=ED m=ECsto Praha;17000;
LABEL;WORK;ENCODING=QUOTED-PRINTABLE;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-
CZ;CHARSET=ISO-8859-2:Kamenick=E1 652/48=0D=0A17000=20Praha=0D=0A

END:VCARD
```

This is probably the best way to have it clear as parseable adress-record when you download every VCARD File and parse the data into a MySQL database.

Please parse all data you see on the webpage into the MySQL database, like tel, fax, email, street, housenr, zip code, city, etc... – just EVERYTHING needs to be parsed into the database!!!

The ZLATE STRANKY Website has 2 areas:

- a) Private customers (White pages directory)
- b) Business Customers (Yellow pages directory)

You also should parse and get all business entries, where it's important, that you get the tel, fax, mobile nr and even the category in which business category the listing is published...

I recommend you to always download the VCF file, but also parse the data on the webpage as well... Just import it even twice like:

Name Webpage:	Meier
Name VCF File:	Meier
Firstname Webpage:	Frank
Firstname VCF File:	Frank

(so in the MySQL Database would be the same information 2 times, where it can be possible, that in the VCF File there are different information than on the webpage... what I hear...)

The business VCF Files entries look like that:

```
BEGIN:VCARD
VERSION:2.1

N;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:MEDIATEL, spol. s r.o.
FN;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:MEDIATEL, spol. s r.o.
TEL;TYPE=WORK;TYPE=VOICE:841 121 212
TEL;TYPE=WORK;TYPE=FAX:224 812 319
URL;TYPE=WORK:http://www.mediatel.cz
EMAIL;INTERNET:info@mediatel.cz
ADR;WORK;ENCODING=QUOTED-PRINTABLE;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2::;Th=E1mova 137/16;Praha;Hlavn=ED m=ECsto Praha;18600;
LABEL;WORK;ENCODING=QUOTED-PRINTABLE;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:Th=E1mova 137/16=0D=0A18600=20Praha=0D=0A

END:VCARD
```

The business entries of companies look like this:

en.zlatastranky.cz/businesses/Praha/H122589/MEDIATEL%2C+spol.+s+r.o./

Auf Facebook tei... Rječnik hrvatsko... http://www.face... Nessus javascript:var%2... Aktuelle Nachr...

franek SEARCH Facebook Twitter Mininova Vuze Blog [16]

Stranky.cz Supersearch Businesses People Maps

What select a category **Where** select a region **In Around**

mediatel 11000 none Find

Keyword, Name, Phone, Heading or Products City, Street, Postcode

Zlaté Stránky > M > Marketing - consultants > M > MEDIATEL, spol. s r.o.

< Back Map Get Directions Save to: My Places Send to: Outlook Email Mobile Share on: My

MEDIATEL, spol. s r.o.
 Thámova 137/16, 18600 Praha 8 - Karlín
 Tel.: 841 121 212
 Fax: 224 812 319
 Category: [Marketing - consultants, Hotels](#)

[www.mediatel.cz](#)
[www.zlatastranky.cz](#)
[www.internet123.cz](#)
[www.inform.cz](#)
[www.travelguide.cz](#)
[info@mediatel.cz](#)

Like Be the first of your friends to like this.

Details Ratings & Reviews

Other Contact Numbers (5)
 Tel.:848 123 123
 Tel.:257 102 102
 Tel.:800 108 001
 Tel.:724 642 624
 Tel.:543 251 704
 Fax: 221 807 447
 View All

Another E-mail Addresses
[jaroslav.vrtiska@travelguide.cz](#)

Description:
 Společnost Mediatel nabízí široké portfolio marketingových nástrojů pro efektivní podporu podnikání. Poskytuje unikátní mix komunikačních kanálů - online platformu [www.zlatastranky.cz](#) s řadou jedinečných funkcí např. 3D návštěvy či video prezentace. Náš Zlaté stránky a m.zlatastranky.cz

See large map

Karlín
 Křížkova
 Sokolovská
 Pernerova
 Magdala

1000 ft
 200 m

e.g. Restaurants Find Nearby

Just be sure that you import all information you can get from the webpage!!! As well as even the Czech language Description of the company as text... and make „View All“ in the script that you get all telephone numbers...

When you build up the database, please first create a own table with all business categories:

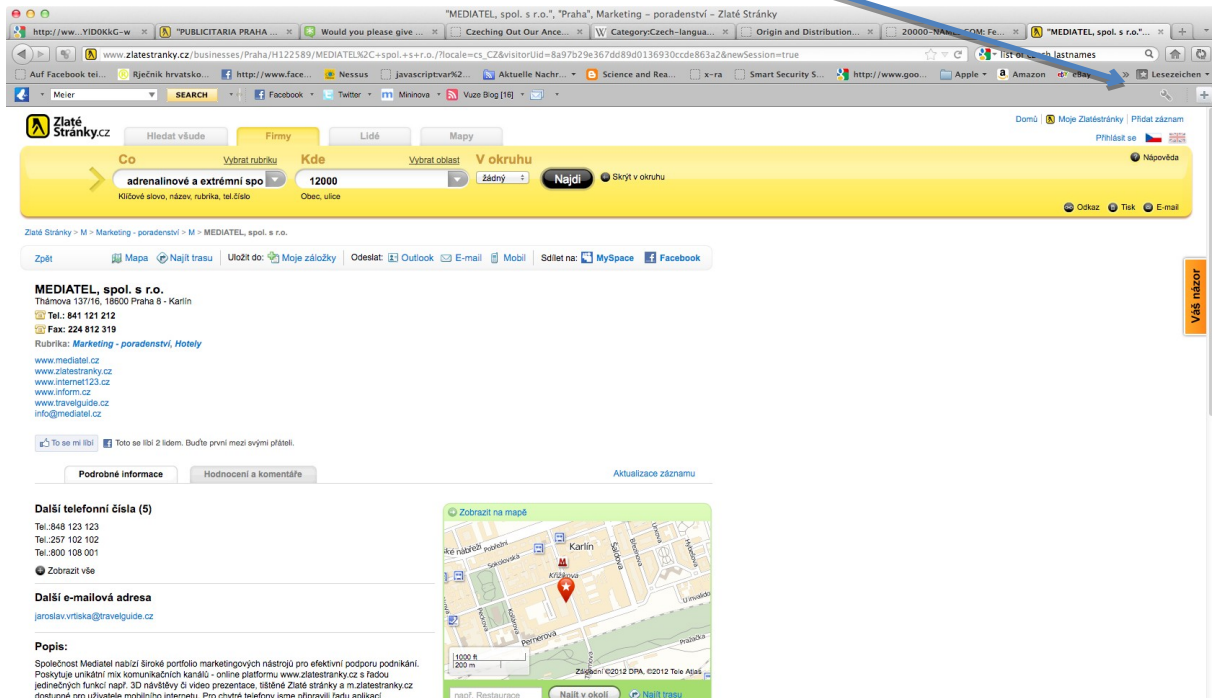
The screenshot shows the Zlaté Stranky.cz website. The browser address bar displays the URL: [en.zlatestranky.cz/businesses/11000/q_abrasive tools/1/](http://en.zlatestranky.cz/businesses/11000/q_abrasive%20tools/1/). The page features a search bar with the text 'mediatel' and a 'SEARCH' button. Below the search bar, there are navigation tabs: 'Supersearch', 'Businesses' (selected), 'People', and 'Maps'. A dropdown menu titled 'select a category' is open, showing a list of categories including 'Abrasive tools', 'Accommodation', 'Accounting', 'Advertising', and 'Advertising - signs & billboards & pos stands'. The main content area displays a business listing for 'BRUSIVO PROBRUS PRAHA' with the address 'Štěrboboholská 1307/44, 10200 Praha 15 - Hostivař'. The listing includes a 'Show telephone' button, a 'Category: Abrasive tools' label, and a website link 'www.brousit.cz'. A map of the area is shown on the right side of the listing. Below the map, there is a section titled 'Also Serving this Area' which lists 'ABRASIV a.s.' with the address 'Pod Borkem 312, 29301 Mladá' and a 'Show telephone' button.

I want that you then crawl all business listings by that order:

- 1.) Select Business Category
- 2.) Select Zip code 11000, 12000, 13000, ... , 99000
And select (find all listings in 75km area...)

Multilanguage Crawling:

I want that you also crawl the Czech-language version of the entry:



TODO:

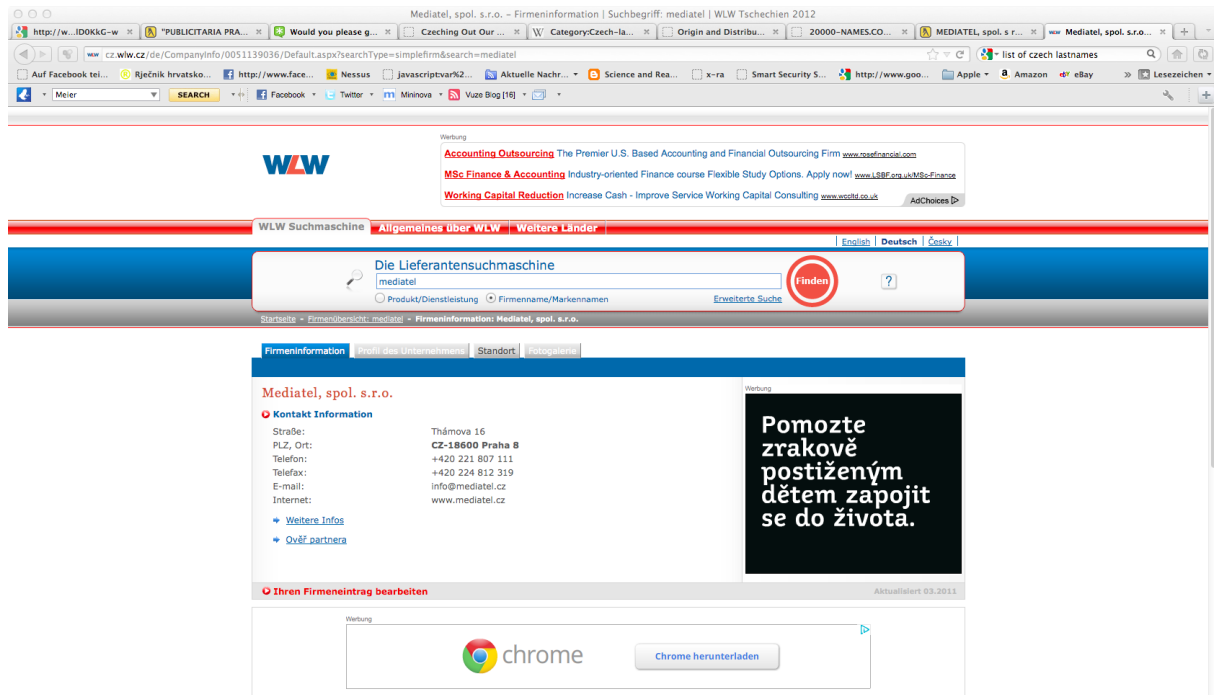
- Please add to the relevant entry the Czech definition of the business category and also the English category
- Please crawl all data as EN (English) selection and as CZ (Czech) selection and add everything you can get from the webpage to the MySQL Database

Information:

- You don't need to download the MAP picture... this we don't need...

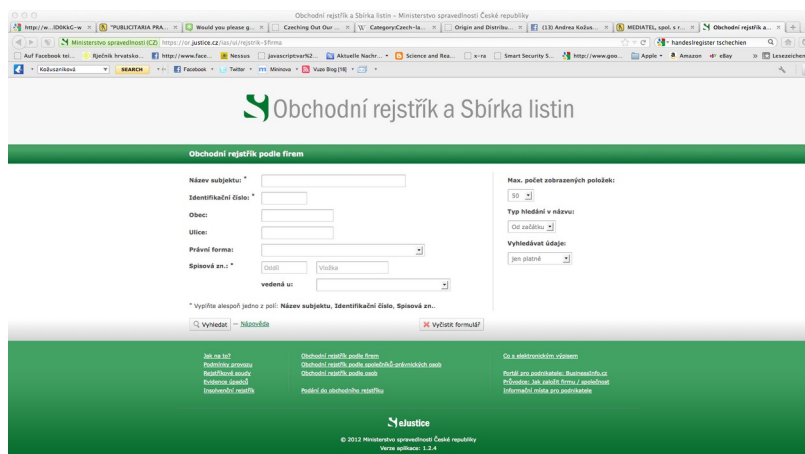
Additional databases:

Because we can also make it better, I want you after you have a list of all companies from Czech Republic and businesses, that you also check every entry on the WLW webpage:



I also want that you retrieve every company in Czech Republic from the government webpage:

[https://or.justice.cz/ias/ui/rejstrik-\\$firma](https://or.justice.cz/ias/ui/rejstrik-$firma)

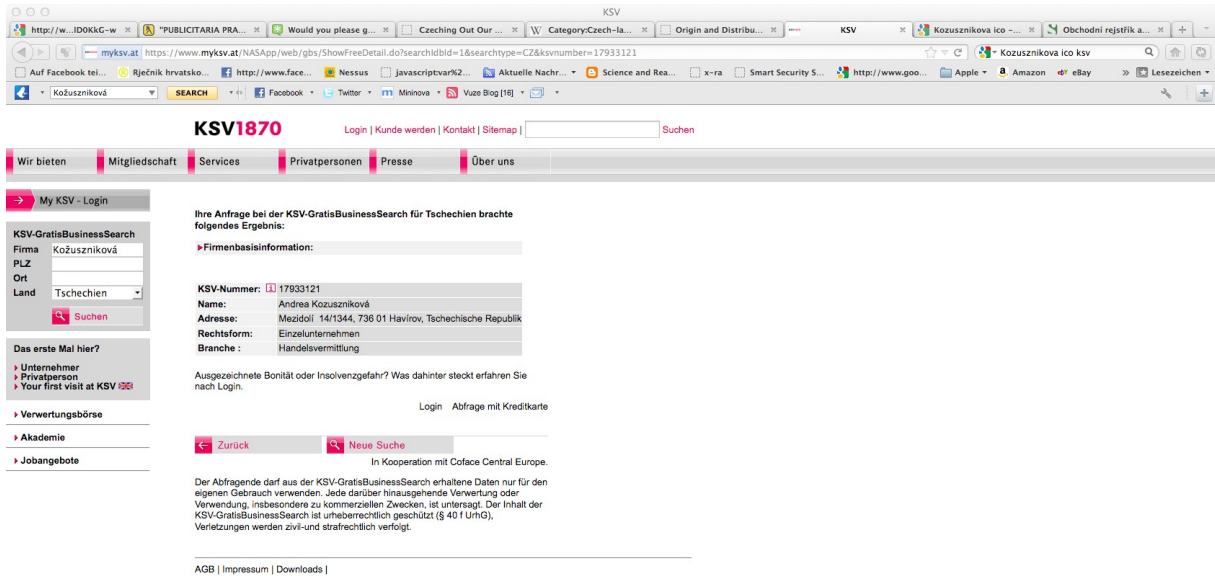


Important to take care of:

- ➔ Czech Republic has special characters, so you have to have them enabled in the database.... like Kožuszniková

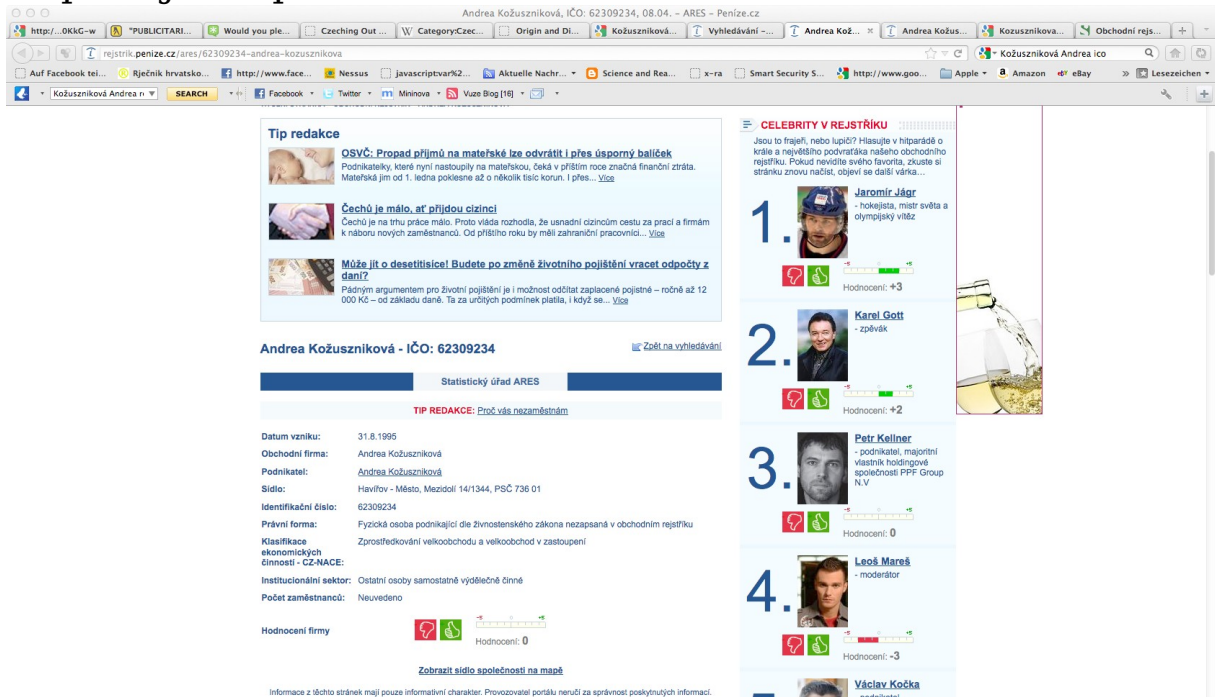
Also parse that database:

The screenshot shows a web browser window with the URL <http://www.ksv.at/KSV/1870/>. The page features a search bar with the text "Kožuszniková" and a "SEARCH" button. Below the search bar, there is a navigation menu with links: "Wir bieten", "Mitgliedschaft", and "Service". A "My KSV - Login" button is also visible. On the left side, there is a search form titled "KSV1870 gratisBusinessSearch" with fields for "Firma", "PLZ", "Ort", and "Land" (set to "Tschechien"). A "Suchen" button is located below these fields. A red banner with the text "Willkommen" is partially visible on the right side.



And : - also parse that webpage database, because here you find also ICO numbers...

<http://rejstrik.penize.cz/andrea-kozusznikova>



Whois

Please make whois lookup for every domain name in Czech Republic you find in the listings. Just make a script that for every domain name (found in email addresses or webpages) the record is retrieved and stored in the database..

```
%whois -h whois.nic.cz mediatel.cz
% (c) 2006-2012 CZ.NIC, z.s.p.o.
%
% Intended use of supplied data and information
%
% Data contained in the domain name register, as well as information
% supplied through public information services of CZ.NIC association,
% are appointed only for purposes connected with Internet network
% administration and operation, or for the purpose of legal or other
% similar proceedings, in process as regards a matter connected
% particularly with holding and using a concrete domain name.
%
% Full text available at:
% http://www.nic.cz/page/306/intended-use-of-supplied-data-and-information/
%
% See also a search service at http://www.nic.cz/whois/
%
%
% Whoisd Server Version: 3.4.0
% Timestamp: Sun Apr 08 21:28:12 2012

domain:      mediatel.cz
registrant:  SB:MEDIATEL
admin-c:     MEDIATEL-ADMIN
nsset:       NSS:MEDIATEL:1
registrar:   REG-1X
status:      paid and in zone
registered:  13.09.1996 02:00:00
changed:     03.04.2012 22:38:35
expire:      28.10.2012

contact:     SB:MEDIATEL
org:         MEDIATEL, spol. s r.o.
name:        MEDIATEL, spol. s r.o.
address:     Thamova 137/16
address:     Praha 8
address:     186 00
address:     CZ
e-mail:      webmaster@zlatestranky.cz
registrar:   REG-CT
created:     10.08.2001 22:13:00
changed:     13.09.2007 14:25:00

contact:     MEDIATEL-ADMIN
name:        Martin Smutny
address:     Thamova 137/16
address:     Praha 8
address:     186 00
address:     CZ
phone:       +420 221807403
e-mail:      msmutny@zlatestranky.cz
registrar:   REG-CT
created:     10.08.2001 22:13:00

nsset:       NSS:MEDIATEL:1
nserver:     ns.zlatestranky.cz
nserver:     au0.zlatestranky.cz
tech-c:      SB:MEDIATEL
tech-c:      MEDIATEL-ADMIN
tech-c:      MEDIATEL-TECH
registrar:   REG-CT
created:     01.10.2007 02:00:00

contact:     MEDIATEL-TECH
name:        Milos Bilwachs
address:     Thamova 137/16
address:     Praha 8
address:     186 00
address:     CZ
phone:       +420 221807240
e-mail:      mbilwachs@zlatestranky.cz
registrar:   REG-CT
created:     10.08.2001 22:13:00
```

Job:

I need somebody who can write a script which parses the webpage with the white-pages (Telephone book) data into a MySQL Database from Zlatestranky.cz and some other business-directories. I need a list of adressdata from people and companies in the Czech Republic, so I want to import from the webpage the phone-book and some other directories.

I have written a detailed instruction what I need as PDF...